SOME ASPECTS OF LINEAR MODEL: BLUS RESIDUALS, DURBIN-WATSON LEMMA, AND TESTABILITY*

by

Somesh Das Gupta
University of Minnesota

Technical Report No. 166

January 1972

University of Minnesota
Minneapolis, Minnesota

*This research was supported by U.S. Army research grant DA-ARO-D-31-124-70-G102.

1. Introduction.

In this paper three aspects of a univariate linear model, namely, (1) optimality of BLUS residuals, (2) Durbin-Watson Lemma, and (3) testability of a linear hypothesis are considered. Although there is no connection among these three problems, they are tackled in this single paper due to the fact that all of them arise from a linear model and either a new or a general proof or a new viewpoint is presented for each of these problems.

Theil [8,9] introduced the concept of BLUS residuals and Grossman and Styan[3] proved some optimal properties of these BLUS residuals. In this paper, a simple and direct proof of the results of Grossman and Styan is given without the restriction imposed by Theil and Grossman and Styan (see (2.13) of Section 2).

A new proof is given for the well-known Durbin-Watson lemma which is used to construct a test of independence of the error variables in a linear model. In the original paper of Durbin and Watson [1] this lemma was not stated correctly and their proof is not clear. Later Anderson [2] gave the correct statement of the lemma. It is shown that this lemma is a simple consequence of Courant-Fisher minimax theorem [5].

For testing a linear hypothesis in a linear model, Roy and Roy [6, 7] introduced different concepts on testability. A geometric explanation of the different situations is presented here and it is also emphasized that the notions of testability introduced by Roy and Roy is somewhat misleading. All the results of Roy and Roy and Millikan [4] are obtained in a simple way without involving unnecessary matrix calculus.

2. On BLUS Residuals.

Some results on BLUS residuals, which are slightly more general than those obtained by Grossman and Styan [3] and Theil [8, 9], are stated with very simple proofs.

Consider a linear model

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{u},$$

where X is a known $n \times q$ matrix of rank q, $\underset{\sim}{\beta}$ is unknown, and $\underset{\sim}{u}: n \times 1$ is a random vector with mean $\underset{\sim}{0}$ and covariance matrix $\sigma^2 I_n$.

A vector of uncorrelated regression residuals is defined by $\underset{\sim}{\gamma}_A = A'\underset{\sim}{Y}$, where $A : n \times m$, and

$$E\underset{\sim}{\gamma}_A = \underset{\sim}{0} \quad \text{and} \quad \mathrm{Cov}(\underset{\sim}{\gamma}_A) = \sigma^2 I_m. \tag{2.1}$$

Let $L : n \times e \; (e = n - q)$ be a matrix such that $L'L = I_e$ and $C(L)$, the vector space spanned by the column vectors of L, is orthogonal to $C(X)$. Note that

$$E(A'Y) = 0 \;\Leftrightarrow\; C(A) \subset C(L).$$

Hence (2.1) is equivalent to $A = LH$ with $H'H = I_m$. When $m = e$, the above condition is equivalent to $AA' = LL'$. This was obtained by Koerts (See [3]) in a lengthy way.

Now, suppose we want to "approximate" $J'\underset{\sim}{u}$ by $A'\underset{\sim}{Y}$ where

$$J'J = I_e, \; A'A = I_e, \; E(A'\underset{\sim}{Y}) = 0 \,. \tag{2.2}$$

The "best" approximation is done by minimizing

$$\mathrm{tr}[\mathrm{Cov}(A'\underset{\sim}{Y} - J'\underset{\sim}{u})]. \tag{2.3}$$

It is clear that if $C(J) = C(L)$ then $J'\underset{\sim}{Y}$ is the best approximation since $J'\underset{\sim}{Y} = J'\underset{\sim}{u}$ in this case.

For general J with conditions (2.2), we proceed as follows.

$$\Sigma_A = \mathrm{Cov}(A'\underset{\sim}{Y} - J'\underset{\sim}{u}) = \mathrm{Cov}[(A-J)'\underset{\sim}{u}], \text{ since } A'X = 0 \tag{2.4}$$
$$= \sigma^2(A-J)'(A-J)$$
$$= \sigma^2(2I_S - A'J - J'A), \text{ using (2).}$$

Thus the problem is to maximize $tr(A'J)$ subject to (2.2) or equivalently

$$A'A = J'J = I_e,\ C(A) = C(L). \tag{2.5}$$

Note that J can be expressed as

$$J = J_1 + J_2, \text{ where } C(J_1) \subset C(L),\ C(J_2) \perp C(L) \tag{2.6}$$
$$= LM + J_2, \text{ where } M : e \times e.$$

Let the rank of M be s. Then

$$M = PDQ', \tag{2.7}$$

where P, Q are $e \times e$ orthogonal matrices,

$$D = \left[\begin{array}{c|c} D_s & 0 \\ \hline 0 & 0 \end{array}\right], \tag{2.8}$$

$D_s : s \times s$ is a diagonal matrix with positive diagonals less than or equal to 1.

We can write $A = LT$ where $T : e \times e$ is an orthogonal matrix. Now,

$$tr(A'J) = tr(T'M) = tr(Q'T'PD) = tr(G'D), \tag{2.9}$$

where $G' = Q'T'P$ is an orthogonal matrix. Since G is orthogonal

$$tr(G'D) \le tr(D) \tag{2.10}$$

and the equality occurs, iff

$$G = \left[\begin{array}{c|c} I_s & 0 \\ \hline 0 & G_1 \end{array}\right] \tag{2.11}$$

where G_1 is an orthogonal matrix. Hence an optimal A is given by

$$LP\left[\begin{array}{c|c} I_s & 0 \\ \hline 0 & G_1 \end{array}\right] Q' \tag{2.12}$$

where P, Q, G_1 are defined as before.

Grossman and Styan [3] and Theil [8, 9] considered the above problem with the following restricted set-up. Let K be an $n \times q$ matrix such that

$$K'J = 0,\ K'X \text{ is nonsingular.} \tag{2.13}$$

This condition is equivalent to

$$C(X) \cap C(J) = \{0\}$$

$$\Leftrightarrow\ C(X)^{\perp} \cap C(J)^{\perp} = \{0\}$$

$$\Leftrightarrow\ J'L \text{ is nonsingular}$$

$$\Leftrightarrow\ M \text{ is nonsingular (i.e., } s = e).$$

So, in this case, the optimum A is given by

$$LPQ'. \tag{2.14}$$

Let A^* correspond to an optimum A in the general case. Next, we shall give a shorter (and direct) proof of an optimality result, more general than that obtained by Grossman and Styan. Their result is (2.19), and, in particular,

$$Ch_{max}(\Sigma_A) \geq Ch_{max}(\Sigma_{A^*}) \tag{2.15}$$

where A satisfies (2.2) and Σ_A is defined by (2.4), and Ch_{max} denotes the maximum characteristic root.

Consider an $e \times 1$ vector $\underset{\sim}{\ell}$ with $\underset{\sim}{\ell}'\underset{\sim}{\ell} = 1$. Note that

$$\underset{\sim}{\ell}'\Sigma_A\underset{\sim}{\ell} = \sigma^2\|A\underset{\sim}{\ell} - J\underset{\sim}{\ell}\|^2,$$

where $\| \ \|$ denotes the standard Euclidean norm. For fixed $\underset{\sim}{\ell}$, $\underset{\sim}{b} \in C(L)$, $\underset{\sim}{b}'\underset{\sim}{b} = 1$,

$$\|\underset{\sim}{b} - J\underset{\sim}{\ell}\|^2$$

attains its minimum value $2 - 2\|J_1\underset{\sim}{\ell}\|$ when $\underset{\sim}{b} = J_1\underset{\sim}{\ell}/\|J_1\underset{\sim}{\ell}\|$, J_1 being defined as in (2.6). This can be seen easily. Thus

$$(1/\sigma^2)\underset{\sim}{\ell}'\Sigma_A\underset{\sim}{\ell} \geq 2 - 2\|J_1\underset{\sim}{\ell}\| = 2 - 2(\underset{\sim}{\ell}'M'M\underset{\sim}{\ell})^{\frac{1}{2}} = 2 - 2(\underset{\sim}{\ell}'QD^2Q'\underset{\sim}{\ell})^{\frac{1}{2}} \ . \qquad (2.16)$$

Taking supremum for both the sides, we get

$$(1/\sigma^2)Ch_{max}(\Sigma_A) \geq 2 - 2\ Ch_{min}(D) = 2\ Ch_{max}(I-D) = (1/\sigma^2)Ch_{max}\ \Sigma_{A^*} \qquad (2.17)$$

since,

$$\Sigma_{A^*} = \sigma^2(A^*-J)'(A^*-J) = \sigma^2[2I_e - 2QDQ']. \qquad (2.18)$$

Note that, if $s < e$, $Ch_{max}(\Sigma_{A^*}) = 2\sigma^2$. This is the result of Grossman and Styan who obtained it in a restricted setup discussed earlier in (2.13).

Using Courant-Fisher minimax theorem [5], the following follows from (2.16) in a straightforward manner:

$$Ch_i(\Sigma_A) \geq Ch_i(\Sigma_{A^*}),\ i = 1,\dots,e \qquad (2.19)$$

where Ch_i denotes the i^{th} largest characteristic root. It may be remarked that for any two $p \times p$ p.s.d. matrices Γ_1 and Γ_2,

$$\Gamma_1 - \Gamma_2 \text{ is p.s.d.} \Rightarrow Ch_i(\Gamma_1) \geq Ch_i(\Gamma_2),\ i = 1,\dots,p \qquad (2.20)$$

although the converse is not necessarily true. It was pointed out by Grossman and Styan that Theil's conjecture "$\Sigma_A - \Sigma_{A^*}$ is p.s.d." is false; however, Theil's conjecture is <u>almost</u> true, in the sense described in the above result (2.19).

3. On the Durbin-Watson Lemma.

For testing independence of the error variables in a linear model, Durbin and Watson [1] suggested a test which hinges on a lemma in matrix algebra known as the Durbin-Watson Lemma. In the paper of Durbin and Watson there are some wrong statements and the proof of the lemma does not appear to be clear although the basic idea is correct. Later Anderson [2] gave the corrected version of the lemma. Here a simple proof of this lemma is given using the Courant-Fischer minimax theorem [5].

Lemma.

Let $\nu_1 \geq \dots \geq \nu_m$ be the characteristic roots of a matrix $Q' \Lambda Q$, where $Q : n \times m$ $(m \leq n)$, $Q'Q = I_m$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Then

$$\lambda_{i+n-m} \leq \nu_i \leq \lambda_i \ , \ i = 1, \dots, m.$$

The proof of this lemma is simple and is omitted.

Let $M = I_n - X(X'X)^{-1}X'$ where $X : n \times k$ is of rank k. There exists a matrix $P : n \times (n-k)$ such that $P'P = I_{n-k}$, $PP' = M$, $P'X = 0$. Let

$$r = \underset{\sim}{u}'(MAM)\underset{\sim}{u} \ / \ \underset{\sim}{u}'M\underset{\sim}{u}$$

where $A : n \times n$ is a symmetric matrix. There exists an orthogonal matrix R and a diagonal matrix $N = \text{diag}(\nu_1, \dots, \nu_{n-k})$, $\nu_1 \geq \dots \geq \nu_{n-k}$ such that $P'AP = R'NR$. Then

$$r = \frac{v'(P'AP)v}{v'v} = \frac{w'Nw}{w'w} = \frac{\sum_1^{n-k} v_i w_i^2}{\sum_1^{n-k} w_i^2} \tag{3.1}$$

where $\underset{\sim}{v} = P'\underset{\sim}{u}$, $\underset{\sim}{w} = R\underset{\sim}{v} = RP'\underset{\sim}{u}$. The spectral decomposition of A is given by

$$A = \sum_{i=1}^{n} \lambda_i \ (\underset{\sim}{\ell}_i \underset{\sim}{\ell}_i')$$

where $L = (\underset{\sim}{\ell}_1 \dots \underset{\sim}{\ell}_n)$ is an orthogonal matrix. Suppose s specified $\underset{\sim}{\ell}_i$'s lie in $C(X)$, the vector space spanned by the columns of X. Then

$$P'AP = \sum_{i=1}^{n-s} \lambda_i^* (P'\ell_i^*)(\underset{\sim}{\ell}_i^{*\prime} P) = P'A^*P$$

where

$$A^* = \sum_{i=1}^{n-s} \lambda_i^* (\underset{\sim}{\ell}_i^* \underset{\sim}{\ell}_i^{*\prime}),$$

and $\underset{\sim}{\ell}_i^*$'s are $\underset{\sim}{\ell}_i$'s other than those which are known to lie in $C(X)$ and λ_i^*'s are the corresponding λ_i's. Suppose $\lambda_1 \geq \dots \geq \lambda_n$, $\lambda_1^* \geq \dots \geq \lambda_{n-s}^*$, $L^* = (\ell_1^* \dots \ell_{n-s}^*)$ $\Lambda^* = \mathrm{diag}(\lambda_1^*, \dots, \lambda_{n-s}^*)$. Then

$$P'AP = P'A^*P = Q'\Lambda^*Q$$

where $Q' = P'L^*$. Note that

$$I_{n-r} = (P'L)(P'L)' = P'LL'P = P'L^*L^{*\prime} P = Q'Q$$

using the above lemma, we get

$$\lambda_{i+k-s}^* \leq \nu_i \leq \lambda_i^* \text{ , } i = 1, \dots, n-k.$$

It follows from the proof of the above lemma that $\nu_i = \lambda_i^*$, $i = 1, \dots, n-k$, iff the characteristic vectors $\ell_{n-k+1}^*, \dots, \ell_{n-s}^*$ lie in $C(X)$; a similar result holds for the other inequality. Thus

$$\frac{\sum_{i=1}^{n-k} \lambda_{i+k-s}^* w_i^2}{\sum_{i=1}^{n-k} w_i^2} \leq r \leq \frac{\sum_{i=1}^{n-k} \lambda_i^* w_i^2}{\sum_{i=1}^{n-k} w_i^2} . \tag{3.2}$$

In general, (when nothing is known about $C(X)$),

$$\frac{\sum_{i=1}^{n-k} \lambda_{i+k-s} w_i^2}{\sum_{i=1}^{n-k} w_i^2} \leq r \leq \frac{\sum_{1}^{n-k} \lambda_i w_i^2}{\sum_{1}^{n-k} w_i^2} . \tag{3.3}$$

This follows from an application of the above lemma to $P'AP$. Durbin-Watson's lemma is precisely (3.1), (3.2), and (3.3).

Suppose, instead of having $\text{Cov}(\underset{\sim}{u}) = \sigma^2 I$ we have $\text{Cov}(\underset{\sim}{u}) = \Sigma$ and Σ has the same characteristic vectors as A. Moreover, assume that $C(X)$ is spanned by k characteristic vectors of Σ. Then we choose the columns of P as the remaining $(n-k)$ characteristic vectors of Σ. In that case, $P'AP = \Delta_1$, a diagonal matrix, and $P'\Sigma P = \Delta_2$, a diagonal matrix. Let

$$\eta = \Delta_2^{-\frac{1}{2}} \underset{\sim}{v}.$$

Then

$$r = \frac{\eta'(\Delta_2^{\frac{1}{2}} \Delta_1 \Delta_2^{\frac{1}{2}})\eta}{\eta' \Delta_2 \eta}$$

and

$$\text{Cov}(\eta) = I_{n-k}.$$

Moreover, if Σ and A are structured suitably, one may obtain bounds on r by varying P (and thus varying Δ_1 and Δ_2).

4. Testability of a Linear Hypothesis.

The standard linear model states the mean $\underset{\sim}{\mu}$ of a random vector $\underset{\sim}{Y} : n \times 1$ as $\underset{\sim}{\mu} = A'\underset{\sim}{\theta}$, where A' is $n \times m$ known matrix of rank r and $\underset{\sim}{\theta} : m \times 1$ is unknown. A linear hypothesis is stated as $H : G'\underset{\sim}{\theta} = \underset{\sim}{0}$ where G is known $s \times m$ matrix (say, of rank s). In standard terminology, H is said to be testable, iff $G'\underset{\sim}{\theta}$ is linearly estimable, i.e., $C(G) \subset C(A)$, where C denotes the vector space spanned by the columns. In two papers [6 , 7] Roy and Roy considered H with general type of G matrix in H and introduced the concepts of complete testability, partial testability and non-testability.

It is the purpose of this note to indicate that these concepts are somewhat misleading and to clarify the situation through geometric interpretations.

The model simply states $\mu \in C(A')$ and $\underset{\sim}{\theta}$ serves as coordinates of μ with respect to some spanning vectors. The correspondence between $\underset{\sim}{\mu}$ and $\underset{\sim}{\theta}$ is not unique unless $r = m$. Our interest is on $\underset{\sim}{\mu}$ and any meaningful hypothesis should be expressed in terms of $\underset{\sim}{\mu}$; $\underset{\sim}{\theta}$ serves as an auxiliary parameter. However, in many situations, it is convenient to express the hypothesis in terms of θ since the components of $\underset{\sim}{\theta}$ may have some simple interpretations (although, artificial) in terms of the experiment considered. The present author believes that any linear hypothesis, such as H above, is meaningful and testable (including the trivial cases) when it is "equivalently" expressed in terms of $\underset{\sim}{\mu}$. This idea will now be clarified in the following (non-standard) development; no proof will be given for the straightforward results given below.

Let T be a linear transformation of E^m to E^n ($\underset{\sim}{\mu} = T\underset{\sim}{\theta}$, $\underset{\sim}{\mu} \in E^n$, $\underset{\sim}{\theta} \in E^m$). The model states that $\underset{\sim}{\mu} \in R(T)$, the range space of T, the domain of T being E^m. Suppose that $\dim R(T) = r$. Let N_T be the null space of T whose dimension is $m - r$. A linear hypothesis states that the domain of T (and, so the range of T) is restricted to a vector subspace. In particular, consider a linear hypothesis $H_0 : D(T) = V$, where D denotes the domain, and V is a vector subspace of E^m of dimension $m - s$. Since our interest is on μ, or, in other words, the range space of T, we shall consider two hypotheses $D(T) = V_1$ and $D(T) = V_2$ to be <u>equivalent</u> iff $TV_1 = TV_2$. The degrees of freedom corresponding to a hypothesis H_0: $D(T) = V$ is defined by $\dim R(T) - \dim T(V)$.

(In the sequel, $\tilde{V}$ will denote the vector space which is the orthogonal complement of V.) It can be seen easily that $TV_1 = TV_2$, iff $V_1 + N_T = V_2 + N_T$ which is the same as saying that the orthogonal projections of V_1 and V_2 on $\tilde{N}_T$ are the same. Note that, with the above definition $D(T) = E^m$ and $D(T) = \tilde{N}_T$ are equivalent.

The inverse image of TV is $V + N_T$ which is the largest vector space whose T-transform is TV. Any hypothesis H_0: $D(T) = V$ has an <u>equivalent</u> representation H_0^*: $D(T) = V + N_T$, although $V + N_T \supset V$. The degree of freedom h of the hypothesis H_0 is

$$\begin{aligned} h &= \dim R(T) - \dim T(V) \\ &= r - [\dim(V + N_T) - \dim N_T] \\ &= m - \dim(V + N_T). \end{aligned}$$

Note also, $h \leq m - \dim V$.

In the standard terminology, H_0: $D(T) = V$ is said to be testable, iff $\tilde{V} \subset \tilde{N}_T$ i.e., $V + N_T = V$, and, in that case, $h = m - \dim V = s$.

Suppose V is such that $V + N_T = E^m$, i.e., $D(T) = V$ is equivalent to $D(T) = E^m$. Such a hypothesis puts no restrictions on the model and its degree of freedom is $m - \dim(V + N_T)$ which equals 0. It is really "equivalent" to the model. This case may be termed as nontestable; but, are we testing θ or testing μ "through" $\underset{\sim}{\theta}$!

Roy and Roy[6,7] classified the testing problems into different categories which may be stated as $V \subset V + N_T \subset E^m$, $V = V + N_T \subset E^m$, $V + N_T = E^m$. These categories are respectively equal to $\tilde{V} \not\subset \tilde{N}_T$ and $\dim TV < r$, $\tilde{V} \subset \tilde{N}_T$ and $\dim TV < r$, $\dim TV = r$ (or, $\tilde{V} \cap \tilde{N}_T = \{\underset{\sim}{0}\}$).

Recently Millikan [4] suggested a new criterion for estimability using the facts:

(a) $\tilde{V} \subset \tilde{N}_T \Leftrightarrow \dim(P_V \tilde{N}_T) = \dim \tilde{N}_T - \dim \tilde{V}$ where P_V denotes the orthogonal projection on V.

(b) For any matrix B, $\mathrm{tr}(BB^-) = \mathrm{rk}(B)$, where B^- denotes the generalized inverse (Penrose) of B.

To see (a) more easily than Millikan's proof, note that

$$\dim(P_V \tilde{N}_T) = \dim \tilde{N}_T - \dim[\tilde{N}_T \cap \tilde{V}]$$

since $\tilde{V}$ is the null space of P_V. Thus $\dim(P_V \tilde{N}_T) = \dim \tilde{N}_T - \dim \tilde{V}$ if, and only if, $\dim \tilde{V} = \dim[\tilde{N}_T \cap \tilde{V}]$ which is equivalent to $\tilde{V} \subset \tilde{N}_T$.

It seems that a more simplified criterion for estimability is to check whether

$$P_{N_T}(\tilde{V}) = \underset{\sim}{0}.$$

This can be stated in terms of relevant matrices and their generalized inverses; as a matter of fact, one has to compute the generalized inverse of only one matrix, namely for computing P_{N_T}.

References

[1] Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression I. Biometrika **37** 409-428.

[2] Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression II: Corrections to Part I. Biometrika **38** 177.

[3] Grossman, S. I. and Styan, G.P.H. (1970). Uncorrelated regression residuals and singular values. The Economics Series Tech. Report No. 32. Stanford University, Stanford, California.

[4] Millikan, G. A. (1971). New criteria for estimability for linear models. Ann. Math. Statist. **42** 1588-1594.

[5] Rao, C. R. (1965). Linear Statistical Inference and its Applications. John Wiley & Sons, Inc., New York.

[6] Roy, S. N. and Roy, J. (1958). A note on testability in normal model I ANOVA and MANOVA. Cal. Math. Soc. Golden Jubilee Commemoration Vol., Part III. Calcutta Math. Soc., Calcutta. Pp. 333-339.

[7] Roy, S. N. and Roy, J. (1959). A note on a class of problems in 'normal' multivariate analysis of variance. Ann. Math. Statist. **30** 577-581.

[8] Theil, Henri (1965). The analysis of disturbances in regression analysis. J. Amer. Stat. Assoc. **60** 1067-1079.

[9] Theil, Henri (1968). A simplification of the BLUS procedure for analyzing regression disturbances. J. Amer. Stat. Assoc. **63** 242-251.